Article:

Quinlan, Philip Thomas orcid.org/0000-0002-8847-6390, Roodenrys, S and Miller, Leonie (2017) Serial reconstruction of order and serial recall in verbal short-term memory. Memory & Cognition. pp. 1-18. ISSN 0090-502X

https://doi.org/10.3758/s13421-017-0719-y

Memory & Cognition

Serial reconstruction of order and serial recall in verbal short-term memory

Journal:	*Memory & Cognition*
Manuscript ID	MC-ORIG-16-320.R2
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	09-May-2017
Complete List of Authors:	Quinlan, Philip; The University of York, Psychology; Roodenrys, Steven; University of Wollongong, Psychology; Miller, Leonie; University of Wollongong, Psychology
Keywords:	Serial recall, reconstruction of order, verbal short-term memory, redintegration

Serial reconstruction of order and serial recall in verbal short-term memory

Philip T. Quinlan

The University of York, UK

Steven Roodenrys and Leonie M. Miller

The University of Wollongong, Australia

Author Note

Please address any correspondence regarding the work to Philip Quinlan, Department of Psychology, The University of York, Heslington York, North Yorkshire, YO10 5DD, UK. Email: philip.quinlan@york.ac.uk. We would like to express our sincere thanks to Alan Baddeley and Graham Hitch for useful discussions on this work.

Abstract

A series of experiments was carried out on verbal short-term memory for lists of words. In the first experiment, participants were tested via immediate serial recall and word frequency and list set size were manipulated. With closed lists the same set of items was repeatedly sampled, with open lists no item was presented more than once. In serial recall, effects of word frequency and set size were found. When a serial reconstruction of order task was used, in a second experiment, robust effects of word frequency emerged but set size failed to show an effect. The effect of word frequency in order reconstruction were further examined in two final experiments. The data from these experiments revealed that the effects of word frequency are robust and are apparently not exclusively indicative of output processes. A multiple mechanisms account is adopted in which word frequency can influence both retrieval and pre-retrieval processes.

.

Serial reconstruction of order and serial recall in verbal short-term memory

One of the most well-known techniques for studying verbal short-term memory is to present a short list of words to a person and then get them to report back the words in the order in which they were presented. This is known as the *immediate serial recall task.* To perform well in the task, participants must remember both the actual words (i.e., the *items*) *and* the order in which the words occurred. Many models of short-term memory, as defined relative to immediate serial recall, incorporate distinct mechanisms for maintaining the two types of information (e.g., Brown, Neath, & Chater, 2007; Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1992; Farrell & Lewandowsky, 2002; Henson, 1998; Page & Norris, 1998) and it has been suggested that different brain regions are involved in dealing with these, respectively (Majerus, 2009). In this vein, attempts have been made to establish whether different factors differentially influence memory for item vs. order information. For example, it has often been suggested that the deleterious effect of having similar sounding words within a list, selectively disrupts memory for the order of the items (e.g., Baddeley, 1986).

In order to align the effects of other variables with the operations of specific mechanisms in models of memory, other tasks have also been used, and the present concerns are with something known as the *order reconstruction task*. In this task, the to-be-remembered (TBR) words that have just been presented are re-presented at test in a new random order, and the participant attempts to reconstruct the order of the words' presentation. A simple assumption is that the order reconstruction task is nothing other than a test of item order information, or, as Whiteman, Nairne and Serra (1994) stated, the reconstruction task "provides a relatively pure index of position memory" (pp. 276-277).

To investigate such a possibility, Whiteman et al. (1994) manipulated word frequency – a variable that reflects item knowledge – in a *free reconstruction of order task*. In this task, participants are free to reconstruct the original sequence in any manner they wish. For example, and unlike in the serial recall task, participants might reconstruct the list in reverse order starting with the last item and then filling in the other items. Whiteman et al. (1994) failed to find an effect of frequency and went on to suggest that word frequency affects item information but not order information.

Several models of serial recall posit that word frequency influences the processes responsible for maintaining item information (e.g. Page & Norris, 1998; Burgess & Hitch, 1992) and thus do not predict an effect of word frequency on measures of order memory. For instance, such models do not predict an effect of word frequency on the number of order errors in serial recall. Although a number of papers have reported such a null effect (e.g. Allen & Hulme, 2006; Miller & Roodenrys, 2012; Poirier & Saint-Aubin, 1996; Stuart & Hulme, 2000), others have found a significant effect of frequency on order information in serial recall (Hulme, Stuart, Brown, & Morin, 2003; Roodenrys, Hulme, Lethbridge, Hinton & Nimmo, 2002).

In addressing this theoretical issue, it is arguable that examination of the effect of word frequency via order reconstruction provides a stronger test of whether word frequency can affect memory for order than does any experiment on serial recall. Consequently, Whiteman et al.'s (1994) null result provides apparently strong evidence against the conclusion that word frequency can affect order memory. Some caution is warranted, however, because we may question whether their experiment reflects on the same memory mechanisms as those underpinning immediate serial recall. Their procedure deviated from the standard immediate serial recall task in two ways; first, it involved free reconstruction rather than a requirement to re-produce the list from beginning to end, and, second, testing was not immediate as there was a 12 second, filled delay between presentation and test.

Other studies have also examined the effect of frequency in reconstruction tasks, but, again, some caution is warranted in considering these. For instance, Thorn, Frankish, and Gathercole (2009) reported data from a paradigm that bears some similarities with the immediate serial recall task. Unfortunately, it is difficult to recover critical details of these experiments because few are provided. Nonetheless, it is reported that words were presented as spoken lists and, at test, the items were re-presented visually. The participant had to say each item in turn in the same order as originally presented. Two datasets are presented and in neither case, is a simple evaluation of the frequency effect given. Although the authors report a diminished frequency effect in reconstruction in comparison to recall, the data as presented appear to show no effect of frequency for reconstruction.

A number of other studies on order reconstruction, have used very slow presentation rates (in comparison to the typical immediate serial recall task) and have included delays between presentation and test (e.g., DeLosh & McDaniel, 1996; Merritt, DeLosh & McDaniel, 2006). In some cases effects of word frequency have emerged (DeLosh & McDaniel, 1996; Merritt, DeLosh & McDaniel, 2006) and in others not (Whiteman et al., 1994). As a consequence, it is very difficult to draw any straightforward conclusions about the effects that word frequency may exert in order reconstruction tasks. It is particularly notable, though, that evidence from paradigms composed of a simple serial presentation of the TBR items followed by an immediate test of memory is lacking.

In the present experiments, the effect of word frequency was examined in both serial recall and reconstruction tasks. A starting point for the experiments is the earlier work by Roodenrys and Quinlan (2000). They examined two variables; (i) word frequency, and, (ii) word set size (or more simply, *set size*), where ‘set size’ refers to how the words within a list were sampled. With *open sampling,* each word was only sampled once in an experiment; with *closed sampling*, words were sampled repeatedly across trials, but no word was repeated within a list. The design of the experiments was based on a factorial combination of open vs. closed lists of high vs. low frequency words. The procedure was serial recall, and across two experiments the findings were clear and consistent. Whereas low frequency words were better recalled from closed vs. open sets there was no effect of set size for the high frequency words. There was also a standard word frequency effect such that high frequency word lists were better recalled than low frequency word lists.

In the same way that we may consider the effects of word frequency across recall and reconstruction tasks, we may also ask how set size effects vary across these tasks. Neath (1997) compared performance with open and closed lists in a variant of the free reconstruction of order task and found that closed lists were more accurately reconstructed than were open lists. Neath’s procedure was much more involved than that typically used in immediate serial recall tasks. Each TBR word was presented for 1 second and before the next item was presented, participants undertook a mental addition problem. Sequences of single digits were presented followed by a target sum and participants had to verify whether the sum of the digits matched the total. This filler addition task lasted for approximately 5 s before the next TBR item occurred. In total the presentation duration of one list took around 37 s and there was a 3 s blank delay before the test was initiated. As before, we may be concerned about the degree to which similar mechanisms are being tapped in this case and in the more standard immediate tests of serial recall. Nonetheless, the nature of the set size effect reported by Neath (1977) is qualitatively the same as that found by Roodenrys and Quinlan (2000) in their tests of serial recall. This at least suggests that set size operates similarly across recall and reconstruction tasks.

The primary aim of the experimental work reported here was to examine word frequency and set size in reconstruction tasks in a bid to understand better the relations between serial recall and reconstruction. In the first experiment, we include a replication of Roodenrys and Quinlan (2000) in which word frequency and set size variables were examined in a serial recall task. The experiment was configured as a web application and a link to the application was disseminated via email. The aim here was to replicate the Roodenrys and Quinlan (2000) findings so as to demonstrate the integrity of data collected over the web. Experiment 2, comprised a serial reconstruction of order task. Again, the data were collected over the web, and, significant effects of word frequency did emerge. This word frequency effect was then explored in two remaining experiments that established the generality and robustness of the phenomenon.

Experiment 1

Method

The experimental tasks were configured using the Qualtrics software in which the corresponding experimental scripts were disseminated via web links. In this way, there was no control over where, or when, participants were tested. The general advice to participants was to undertake the experiment in a quiet place away from distractions. Experiment 1 comprised a replication of Roodenrys and Quinlan (2000), and provides a proof of concept that the novel web-based testing does produce data that are robust and are systematically similar to previous findings.

Design.

Ninety-six high frequency and 96 low frequency words were selected as the stimulus materials and were controlled on several variables known to influence recall performance (taken from Miller & Roodenrys, 2012, see Appendix A). Memory for high frequency and low frequency words was tested in separate testing sessions for every participant. Every participant was tested over two sessions and the order of the word frequency sessions was counterbalanced across the participants. In each session 16 lists of 6 words were presented. There were 16 high frequency and 16 low frequency lists. If the testing involved open lists, then the items were selected without replacement across the lists. Hence a given participant would see every item but only once across all the lists. If the testing involved closed lists, then at the start of the testing six items were randomly selected from the 96 and these were repeatedly sampled across all lists. The order of presentation of the words was determined randomly prior to the start of a trial.

Participants either completed testing sessions comprising open lists or comprising closed lists. Hence set size was a between-participants factor with - Group 1 – open, and Group 2 - closed. More particularly, there were four sub-groups in total as the order of frequency testing was balanced across Groups 1 and 2. Word frequency was a two-leveled factor (high vs. low) and list position was a six-leveled factor.

Participants.

The experiment was run as an online (web-based) laboratory practical and members of the second-year Psychology degree at The University of York were tested as participants. Data from only native English speaking students were examined further and for several participants it was clear that they had failed to follow the instructions properly. For instance, they tested themselves twice with the high frequency lists instead of running separate sessions for the high and low frequency lists. Consequently, the eventual data sets comprised 33 Group 1 - open participants, and, 41 Group 2 - closed participants.

Procedure.

Once the web link was launched a preliminary information screen was presented giving generic instructions about the kind of memory task to be undertaken. Next, they were asked to provide explicit consent to testing. Finally, participants were asked in indicate their age, nationality and to input their unique participant number. The participant number determined whether open or closed lists would be used. The overall cohort of students was divided roughly so that each of the groups was assigned the same *n*. Finally, a start screen was presented which instructed the participant to initiate the experiment when ready.

At the start of each trial a screen displaying the message “Next list” was presented for 1 s. Then the sequence of six TBR words unfolded. Each word was visually presented for 1s and once the list terminated there was a 1 s blank delay before list testing began. Next a screen containing a text box was presented with the message “Input item n” where n (i.e., 1 – 6) indicated the item position currently being requested. The participant entered their response and moved onto the next screen by pressing <Return>. Blank entries were possible by simply pressing <Return>. There was no time constraint on how long a participant could take to enter a response. Six such screens were presented in sequence. A final <Return> key press moved the experiment onto the next trial.

Results and Discussion

All responses were scored strictly so that an item had to be reported correctly in its correct position. As a consequence, each list position was scored as a total correct over the 16 lists and expressed as a proportion. (Descriptive statistics for all conditions of interest in all of the experiments are provided in tabular form in Appendix B.)

Figure 1 provides a graphical illustration of the summary data for the conditions of interest. The data were entered into a split-plot ANOVA in which word frequency (high vs. low), and item position (1-6) were entered as fixed repeated measures factors. Set size (open vs. closed) was entered as a fixed between-participants factor and participants was entered as a random factor. This analysis revealed statistically significant main effects of both word frequency, $F(1, 72) = 57.98$, $p < .001$, partial $\eta^2 = .446$, and, item position, $F(5, 360) = 150.47$, $p < .001$, partial $\eta^2 = .676$. The main effect of set size failed to reach statistical reliability, $F(1, 72) = 1.97$, $p = .164$, partial $\eta^2 = .027$.

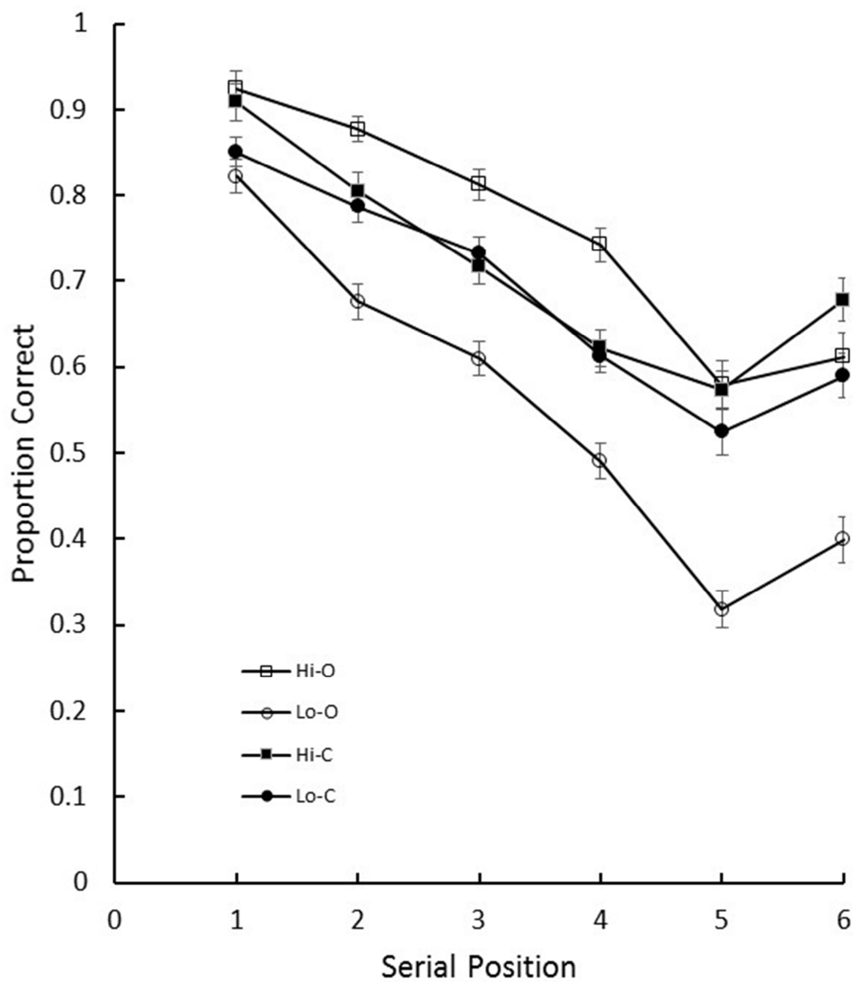

Figure 1. Serial position curves for the four serial recall conditions of interest in Experiment 1. Error bars reflect 1 SE of the particular condition mean. In the notation used, Hi-O stands for high frequency words open set lists, Hi-C stands for high frequency words closed set lists, Lo-O stands for low frequency words open set lists, and Lo-C stands for low frequency words closed set lists.

All of the two-way interactions also reached statistical significance, $F(1, 72) = 29.37$, $p < .001$, partial $\eta^2 = .290$, for the word frequency x set size interaction; $F(5, 360) = 5.68$, $p < .001$, partial $\eta^2 = .073$, for the item position x set size interaction; and, $F(5, 360) = 2.83$, $p < .05$, partial $\eta^2 = .038$, for the word frequency x item position interaction. Finally, the word frequency x item position x set size interaction was also statistically reliable, $F(5, 360) = 12.87$, $p = .001$, partial $\eta^2 = .055$.

These patterns of statistical difference replicate those reported by Roodenrys and Quinlan (2000). Although further comparisons are possible, for the sake of brevity and, as the focus of the current research is on the reconstruction task, no further subsidiary analyses of the recall data are reported. Given that the data replicate those previously reported by Roodenrys and Quinlan (2000), we have shown that the novel web-based testing conditions are reliable and, therefore, that we can approach the data from the ensuing order reconstruction task with some confidence.

Experiment 2

In Experiment 2 a serial order reconstruction task was used. The choice of the particular task was in part influenced by consideration of some recent results reported by Clarkson, Roodenrys, Miller and Hulme (2016). Clarkson et al. (2016) examined performance in a paradigm, in which on each trial and following the serial presentation of a list of words, the words were re-presented in a new random order. Participants had to write down the first word as presented, then the second, then the third and so on until they had attempted to re-create the original presentation order of the words.

Across a series of experiments, Clarkson et al. (2016) examined the effect of phonological neighborhood size. For instance, in their Experiment 2 word lists comprised words solely from either *dense neighborhoods* or *sparse neighborhoods*. A word from a dense neighborhood shares sets of phonemes with many other words whereas a word from a sparse neighborhood shares them with few words. In this way, phonological neighborhood size is a variable that reflects long-term knowledge about how words are phonologically related to one another.

On the assumption that order reconstruction does not rely on accessing such item information because the items themselves are represented at test, then there is no reason to predict that the phonological neighborhood size should influence performance. However, Clarkson et al. (2016) found that participants were more accurate in reconstructing list of items taken from dense neighborhoods than they were in reconstructing lists of items taken from sparse neighborhoods. Without attempting to provide a detailed explanation of this pattern of results here, it suffices to note that Clarkson et al. (2016) have provided more evidence of how the process of order reconstruction is affected by item information despite the items being represented at test (cf. Neath, 1977).

On the grounds that serial order of reconstruction has been shown to be effective in exposing effects of item information in immediate verbal memory, the intention here was to use this in the next experiment. As discussed, the data from order reconstruction tasks concerning effects of word frequency is mixed and, in some cases, it is difficult if not impossible to recover what the exact nature of the testing conditions were. Here we examined effects of word frequency and set size when participants were forced, in an immediate test, to reconstruct the order of the items in the same sequence as at presentation. An intention was to examine performance under conditions that closely matched those typically used in immediate serial recall. Indeed, the identical presentation conditions to those used in Experiment 1 were used in Experiment 2. In this way, we can begin to construct a clearer picture of the differences in memory when testing is either by serial recall or serial reconstruction.

Method

In Experiment 2 performance in a serial reconstruction of order was examined with the materials used in Experiment 1. Again, the Qualtrics software was adapted to run the experiment and testing was carried out over the web in the manner described for Experiment 1. At test, the original words were re-presented as a single list on the screen. The order of the items on the screen was determined on a random basis at the start of each trial and differed from the presentation order. The page also contained the message "Input item n" where n (i.e., 1 – 6) and a text box for typing a word. Participants were therefore forced to type in their responses in sequential order. Once they typed a response pressing <Return> moved the experiment onto the next screen. This screen retained the word list (as just described) and instructions to make their next response.

The materials were the same as in Experiment 1 and the design mirrored that for Experiment 1. Participants were tested twice – once with high frequency word lists and once with low frequency word lists – and the order of testing was counterbalanced across participants. Although the primary aim was to examine the influence of word frequency on order reconstruction, the design also incorporated the set size manipulation from Experiment 1. In this way, both variables were examined in concert. As before, therefore, set size (open vs. closed) was tested as a between-participants factor and the word frequency and item position factors were as used in Experiment 1.

Participants

Participants were sourced from the cohort of first year undergraduates at the University of York. A blanket email was sent to the cohort and the volunteer participants were enrolled on agreeing to take part. All participants were provided with course credit for taking part. Only native English speakers were tested and the sample size was determined by the number of respondents that signed up and agreed to be tested in a two-week period prior to the end of the Spring term. Although 57 participants were tested, due to a variety of web glitches, 50 data sets were analyzed in total. Group 1 (the open group) comprised 23 participants and Group 2 (the closed group) comprised 27 participants.

Results and Discussion

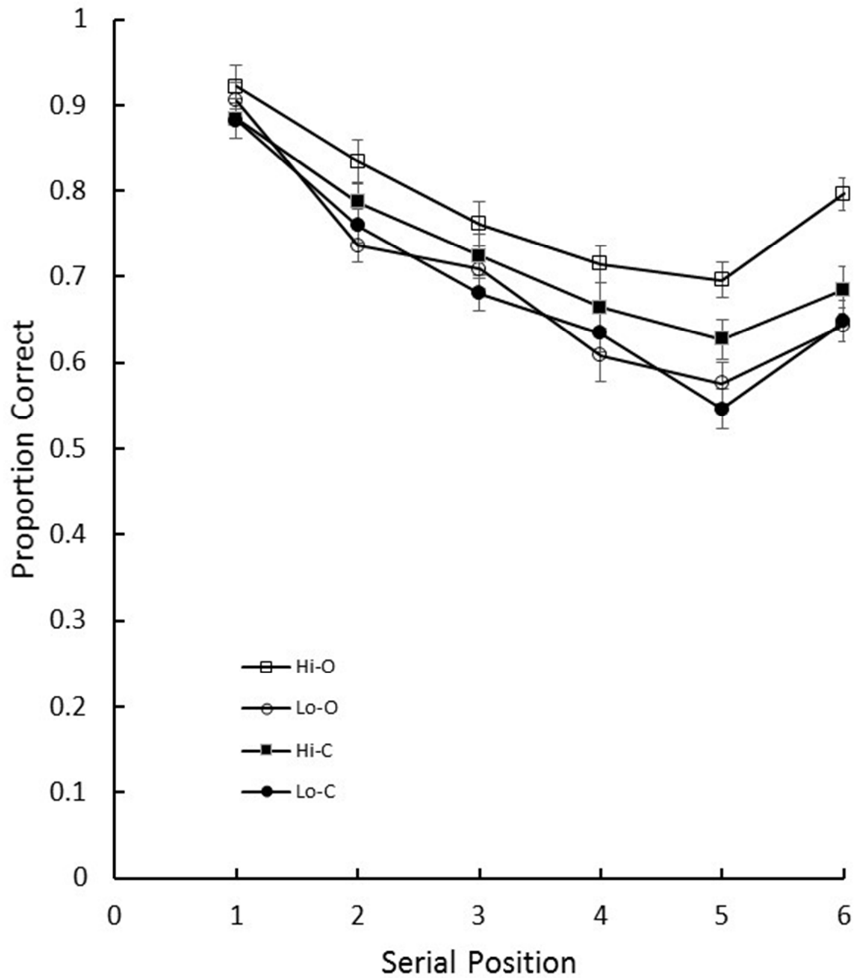

Figure 2. Serial position curves for the four serial reconstruction conditions of interest in Experiment 2. Error bars reflect 1 SE of the particular condition mean. In the notation used, Hi-O stands for high frequency words open set lists, Hi-C stands for high frequency words closed set lists, Lo-O stands for low frequency words open set lists, and Lo-C stands for low frequency words closed set lists.

A graphical illustration of the summary data of interest is provided in Figure 2. The data were scored and analyzed in the same manner as in Experiment 1. This analysis revealed statistically significant main effects of both word frequency, $F(1, 48) = 14.94$, $p < .001$, partial $\eta^2 = .231$, and, item position, $F(5, 240) = 69.23$, $p < .001$, partial $\eta^2 = .591$. The main effect of set size failed to reach statistical reliability, $F < 1.0$. The only other test to reach statistical significance was of the word frequency x item position interaction, $F(5, 240) = 2.88$, $p < .05$, partial $\eta^2 = .057$; $F(1, 48) = 2.53$, $p = .118$, partial $\eta^2 = .050$, for the word frequency x set size interaction; $F < 1.0$, for the item position x set size interaction; and, $F(5, 240) = 1.10$, $p = .356$, partial $\eta^2 = .022$, for the three-way, frequency x item position x set size interaction. So, in sum, the data reveal no effects of set size but robust effects of word frequency and item position.

Of prime interest is that when memory was tested via serial reconstruction of order, effects of word frequency occurred. Participants were better in reconstructing the order of high frequency word lists than low frequency word lists. On these grounds, the data sit comfortably with the view that participants 'do remember and use item information to compete the test' (Neath, 1997, p. 262). This particular finding is followed up in the remaining experiments.

Aside from this positive result, there is the null result regarding set size. The data failed to show any systematic difference in performance across the open and closed lists. This null finding contrasts with the findings in the serial recall task (see Roodenrys & Quinlan, 2000, and Experiment 1 here) and with the robust effects reported by Neath (1997). In order to examine the null effect in more detail the data were examined further using Bayesian methods (Rouder, Morey, Speckman &

Province, 2012). From Roodenrys and Quinlan (2000) and the serial recall task reported above, the effects of set size emerged most strongly and consistently with low frequency but not high frequency words. This set size effect is not present in the data from Experiment 2. To examine this null effect further the data from the low frequency lists were examined in a Bayesian repeated measures ANOVA (JASP Team, 2017) in which the factors were list position and set size as defined for the traditional ANOVA just reported.

Relative to the null model (in which equality is assumed across all cells in the design) a model comprising just the set size factor generated an inverse Bayes factor of 0.280. The null model provided an adequate account of the data and a model including the set size factor failed to provide a statistically better fit, that is, there was substantial evidence in favor of the null model relative to a model including the set size factor (Wetzels, Matzke, Lee, Rouder, Iverson, & Wagenmakers, 2011). As a consequence, we conclude that the data provide *substantial* evidence that there was no effect of set size in the data for the low frequency lists in the serial reconstruction of order task.

In discussing performance in their serial recall task, Roodenrys and Quinlan (2000) argued that the set size variable reflected repeated access of particular lexical entries for items in closed but not in open sets. Critically they assumed that the serial recall task in important respects reflects the operation of long-term memory priming. That is, the repeated presentation of an item primes its lexical entry such that it becomes more readily accessed. Such priming is likely to benefit low frequency lexical entries more than corresponding high frequency lexical entries. High frequency lexical entries are generally more readily accessed then low frequency entries and this may be explained in terms of different resting levels of activation

(Morton, 1969). On these grounds, low frequency lexical entries benefit more from temporary priming than do high frequency items because it makes them more competitive at recall, effectively limiting the set of alternative items that could be recalled in all positions.

The lack of an effect of set size in the reconstruction task in Experiments 2 is consistent with the explanation of the set size effect in serial recall offered above as the reconstruction task limits the set of possible responses in each position to only those presented in the list. However, it stands in contrast to the results of Neath (1997), who did find an effect of set size in a *free* reconstruction of order task.

Aside from differences in the actual paradigms used in the two cases, the kinds of items used in the two cases are very different. Neath (1977) used closed lists comprising items repeatedly sampled from color names ('black', 'blue', 'brown', 'green', 'red', 'white') while the open lists comprised unrelated words. This is important given the evidence that lists containing categorically related items are generally better remembered than are lists containing unrelated items (Poirier & Saint-Aubin, 1995; Wetherick, 1975). In this regard, it seems that the closed vs. open list manipulation used by Neath (1997) is confounded with semantic relatedness which may be producing the pattern in his data. This possibility is given added plausibility by data from another of his experiments in which he found that reconstruction performance was sensitive to word concreteness - a "semantic" variable that reflects item knowledge.

Of course, such speculations do provide pointers to future research, but given the lack of a set size effect in the reconstruction task reported here, this variable is not examined further. It appears that the priming effect of repeatedly presenting a small set of items is nullified by the re-presentation of items in the response phase. The finding of a frequency effect under the same circumstances suggests the two effects reflect the operation of different processes. Indeed, having established a word frequency effect in a serial reconstruction of order task in Experiment 2, the final two experiments build on this finding and establish its robustness and generality. In both experiments a variant on the serial reconstruction of order test was used. Experiment 3 reports on a basic replication of the effect whereas in Experiment 4 its robustness is tested in a more complex design.

Experiment 3

In Experiment 3 the generality of the word frequency effect is explored when participants experienced both high frequency and low frequency word lists at presentation in an intermixed fashion. In Experiment 2, participants were consistently presented with high or low frequency word lists within a given block of trials. We have shown that under these conditions, effects of word frequency in order reconstruction do obtain. However, this experiment was prompted by the findings of Merritt et al. (2006) that the effect in reconstruction of order only occurs when participants experienced lists blocked by condition at presentation and not when they were intermixed. We tested this in the next experiment in which high and low frequency word lists were randomly intermixed.

Methods

Participants.

Participants were 138 third-year undergraduate students in the University of Wollongong Psychology program delivered in Singapore, who took part in the experiment as part of a class exercise. Almost all students in Singapore are bilingual but complete all their schooling in English.

Stimuli and design.

The stimuli used in this experiment were taken from Miller, Roodenrys and Arcioni (under review). Two sets of 30 CVC (consonant-vowel-consonant) stimuli were constructed so that each contained word pairs with the same consonants (see Appendix A). Vowels were selected to create high frequency and low frequency stimuli, respectively. For example, the high frequency stimuli 'mouth', 'team', 'gun', 'book', 'head', 'wife' were counterparts to the low frequency stimuli 'moth', 'tame', 'gown', 'beak', 'hood', 'wharf'. High frequency stimuli ($M = 198.33$, $SD = 124.45$) had reliably greater frequency than low frequency stimuli ($M = 4.80$, $SD = 3.92$), Mann-Whitney $U = 0.00$, $z = -6.66$, $p < .001$. These word sets were also controlled on concreteness, Mann-Whitney $U = 427.50$, $z = -0.33$, $p = .739$, biphone frequency, Mann-Whitney $U = 441.50$, $z = -0.13$, $p = .900$, phonological neighborhood, Mann-Whitney $U = 445.50$, $z = -0.07$, $p = .947$, and vowel length (short, long, and diphthong), $\chi^2 = 2.66$, $p = .265$.

Procedure.

The experimental session was run during tutorial classes and was controlled by a bespoke computer program. Participants were given instructions regarding the serial reconstruction task; for each trial, they would see a fixation cross on the computer screen and after 1000ms the list would be presented, one item at a time in the center of the screen in 24pt Times New Roman font for 1000ms each. After the final item was presented, all items would appear in a random order on the screen from left to right. Using the computer mouse participants were asked to check off the items in the order they appeared in the list by selecting a box adjacent to each word. Once an item was selected a participant could not unselect it and all items had to be selected to complete a trial. The experiment was self-paced with the participant initiating each trial by pressing a 'Next' button on the computer screen.

Prior to the testing phase participants performed 2 practice trials. Each experimental session comprised thirty, 6-item lists of each frequency condition. For each participant, items were randomly arranged into lists; each item appeared across a list set 6 times, but no item appeared in a list more than once. The order of high- and low-frequency lists was randomized in a single block.

Results and Discussion

The data were entered into 2 x 6 repeated measures ANOVA in which word frequency (high vs. low) and item position (1-6) were entered as fixed factors and participants acted as a random factor. The analysis was clear in showing statistically reliable main effects of both word frequency, $F(1, 137) = 205.38$, $p < .001$, partial $\eta^2 = .600$, and, item position, $F(5, 685) = 255.24$, $p < .001$, partial $\eta^2 = .651$, and a statistically significant word frequency x item position interaction, $F(5, 685) = 21.04$, $p < .001$, partial $\eta^2 = .133$. Figure 3 provides a graphical illustration of this pattern of effects.

The data quite clearly reveal a robust effect of word frequency in a serial reconstruction of order task. It is of particular note that, in this case, the effect emerged when pure lists of high and low frequency were randomly intermixed at presentation. In this respect, the data are more indicative of the importance of the constraints at testing rather than those at presentation: the effect of word frequency has emerged under conditions of both blocked (in Experiment 2) and intermixed (in Experiment 3) list presentation conditions.

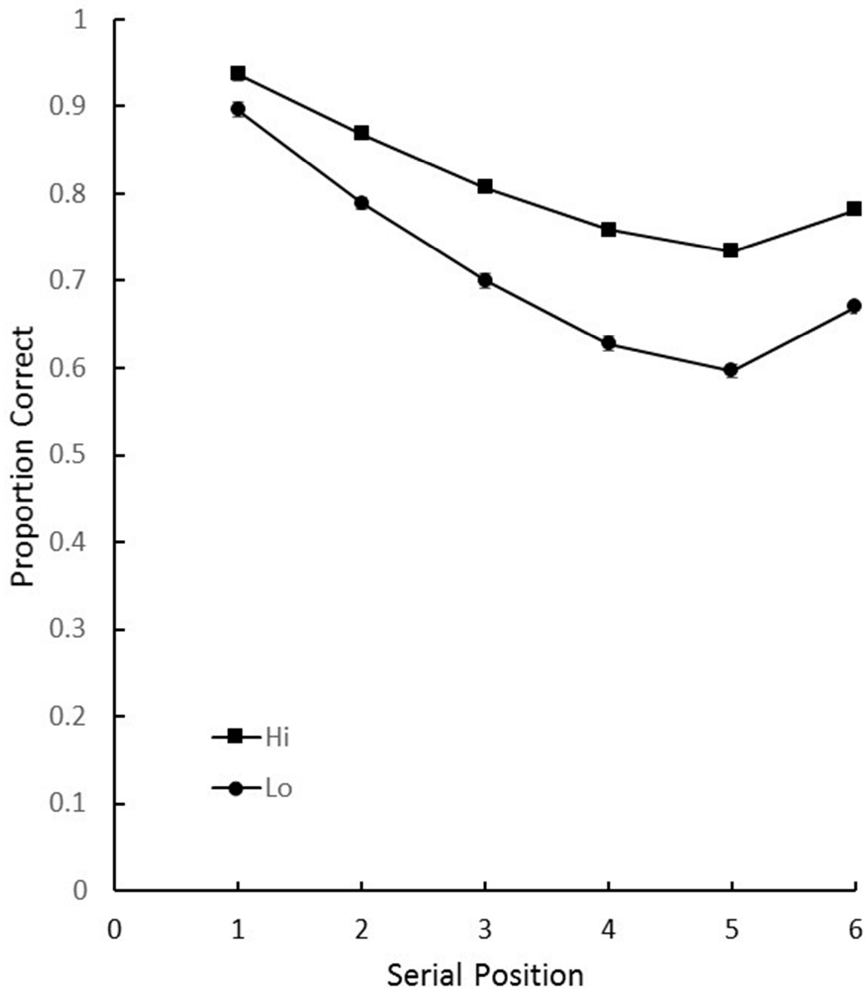

Figure 3. Serial position curves for the two serial reconstruction conditions of interest in Experiment 3. Error bars reflect 1 SE of the particular condition mean. In the notation used, Hi stands for high frequency word lists, and Lo stands for low frequency word lists.

Experiment 4

In our final experiment, we again adopted tests of serial reconstruction of order (i.e., the testing method was as used in Experiment 3), but now we combined these with tests of serial recall. Following the presentation of a list of words, memory was tested either by serial recall or serial reconstruction of order. The type of test was randomized across trials so that participants were unaware of which kind of test would be administered until the time of the test itself. Merritt et al. (2006) reasoned that this kind of procedure guarantees that the manner in which the participants encode the items is the same for both forms of testing. As a consequence, any contingent difference in performance across the two tasks cannot, therefore, be due to differences in item encoding. As in Experiment 3, list type (high- vs low-frequency) was randomly intermixed at presentation.

Methods

Participants.

Participants were 108 third-year undergraduate students the University of Wollongong Psychology program delivered in Singapore who completed the experiment as a subject requirement. The data from 17 students were omitted from the final data set because the pattern of their responses indicated that they were not attempting to recall the lists in full and so were not starting recall at the first item.

Procedure.

The stimuli were identical to those used in Experiment 3 and the procedure was very similar to that for Experiment 3 except that participants now performed 15 trials each of serial recall and serial reconstruction for both high- and low- frequency words. The presentation of items replicated the procedure in Experiment 3. However, recall or reconstruction was post-cued by the response screen that appeared after the presentation of the final item in the list. Reconstruction trials were completed as per Experiment 3. Serial recall trials were prompted by a screen comprising 6 horizontally aligned text boxes that constrained participants to type their responses according to strict serial recall protocol. Practice trials in Experiment 4 included both reconstruction and recall forms to familiarize participants with each type of response.

Results and Discussion

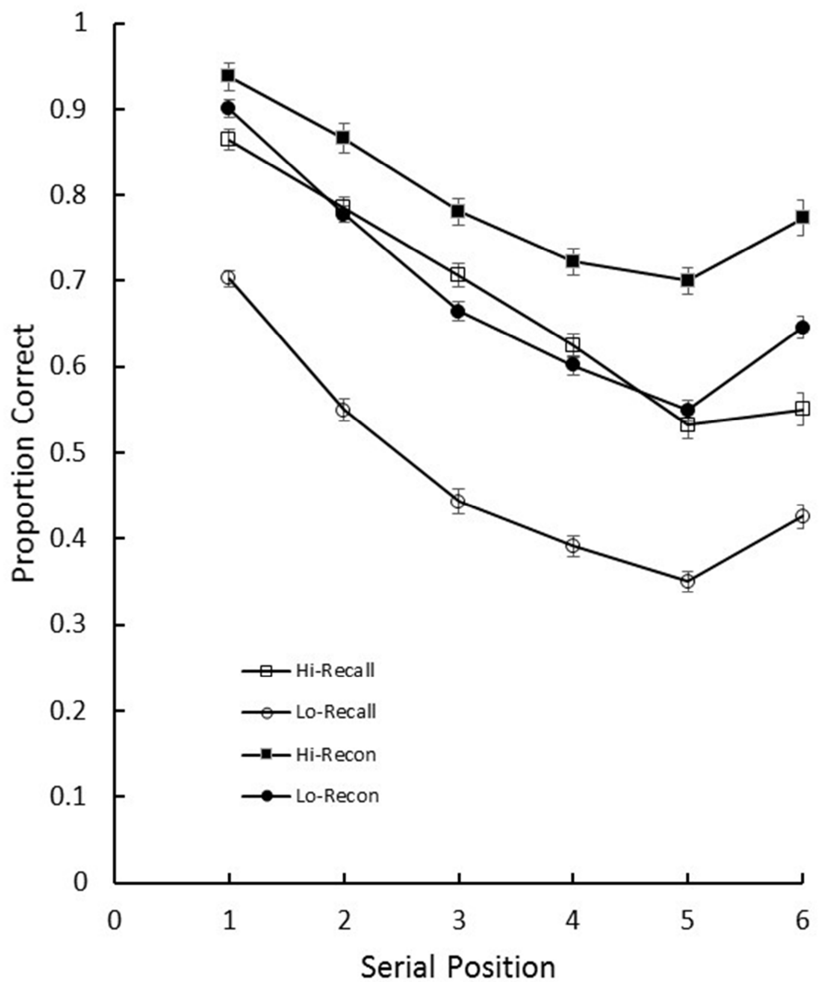

Figure 4. Serial position curves for the four conditions of interest in Experiment 4. Error bars reflect 1 SE of the particular condition mean. In the notation used, Hi-Recall stands for high frequency word lists tested via serial recall, Hi-Recon stands for high frequency word lists tested via serial reconstruction of order, Lo-Recall stands for low frequency word lists tested via serial recall, Lo-Recon stands for low frequency word lists tested via serial reconstruction of order.

A graphical illustration of the summary data for the conditions of interest is provided in Figure 4. The data were entered into 2 x 2 x 6 repeated measures ANOVA in which type of task (recall vs. reconstruction), word frequency (high vs. low) and item position (1-6) were entered as fixed factors and participants acted as a random factor. The analysis revealed that all tests were statistically reliable: $F(1, 90) = 345.63$, $p < .001$, partial $\eta^2 = .793$, for main effect of type of task; $F(1, 90) = 421.46$, $p < .001$, partial $\eta^2 = .824$, for main effect of word frequency; $F(5, 450) = 175.74$, $p < .001$, partial $\eta^2 = .661$, for main effect of item position; $F(1, 90) = 55.70$, $p < .001$, partial $\eta^2 = .382$, for the type of task x word frequency interaction; $F(5, 450) = 5.43$, $p < .001$, partial $\eta^2 = .057$, for the type of task x item position interaction; $F(5, 450) = 9.26$, $p < .001$, partial $\eta^2 = .093$, for the word frequency x item position interaction; and, $F(5, 450) = 8.39$, $p < .001$, partial $\eta^2 = .085$, for the type of task x word frequency x item position interaction.

The statistically significant three-way interaction shows that the effect of word frequency was expressed differently across the item positions in the recall and reconstruction tasks. Subsidiary analyses were carried out via simple main effects tests of the word frequency effect at each of the item positions across the two tasks. These revealed that the effect of word frequency was smaller in the data for the reconstruction task than the recall task in the earlier serial positions. Given the relatively high level of performance in the reconstruction task and the reduced variability in the data in these earlier positions, it seems likely that the three-way interaction is due to the influence of a ceiling effect in earlier positions. The frequency effect in the serial order reconstruction data is therefore remarkably similar to that in the recall data.

Overall, the data are clear in showing robust effects of word frequency in both the recall and reconstruction tasks. The data also show that the effects of word frequency in the serial reconstruction of order tasks reported here (i.e., Experiments 2 and 3) cannot be attributed to some form of special item encoding that comes about when participants know that order memory will be tested in the upcoming trial. Indeed, the nature of the word frequency effect in the reconstruction tasks in Experiments 3 and 4 are remarkably similar. When expressed in terms of Cohen's *d*, the word frequency effect size in Experiment 3 is 1.218 (95% CI spans 1.419 to 1.017; Kirby & Gerlanc, 2013) and in Experiment 4 it is 1.096, (95% CI spans 1.342 to 0.858). This provides some grounds for concluding that the effect of word frequency was essentially the same across these experiments despite the difference in the nature of the list presentation conditions.

General Discussion

We have reported four experiments concerning performance in verbal short-term memory tasks. Across these experiments, we have been able to establish clear effects of word frequency in both serial recall and serial reconstruction of order tasks. Critically, we have been able to establish that the word frequency effects in reconstruction tasks are readily apparent when serial order of reconstruction is tested across a range of different presentation conditions. We have also been able to show differential effects of stimulus set size depending on the task requirements. Set size effects are readily apparent in the data for serial recall tasks but are seemingly absent when reconstruction of order is tested.

Experiment 1 comprised a simple replication of the serial recall experiments reported by Roodenrys and Quinlan (2000). The variables of primary interest were word frequency and set size. The data from the serial recall task replicate those reported by Roodenrys and Quinlan (2000), who argued that the particular pattern of findings is readily accounted for by processes concerning redintegration at test. According to such an account, performance depends on attempting to reinstate the transitory short-term memory traces of the TBR items with corresponding long-term memory representations. Given that retrieval from long-term memory is easier for high frequency than low frequency items (Freedman & Loftus, 1971), the idea is that redintegration is correspondingly more efficient for sequences of common vs. rare words. Although more recent research has determined that the redintegrative potential of a word within a list appears to be a function of list composition (Stuart & Hulme, 2000; Hulme et al., 2003) and item arrangement (Miller & Roodenrys, 2012) rather than specific to the individual item, the same argument holds.

Roodenrys and Quinlan (2000) also argued that the effect of set size was only apparent in the data for the low frequency words because, in a sense, the long-term advantage for high vs. low frequency words far outweighed any additional benefit due to repeating items across the lists. The set size effect was, therefore, only present in the data for the low frequency words. Regardless of these details, the general point is that both variables are seen to influence processes operating at *output* rather than during an earlier stage of processing. As discussion proceeds, we question whether this account provides a useful framework for thinking about the current data.

In Experiment 2, when simple serial presentation was followed by immediate serial reconstruction, a robust effect of word frequency was found. This finding was followed up in Experiments 3 and 4 which involved serial reconstruction of order with different constraints on how participants might encode the lists. In Experiment 3 participants performed only the reconstruction task so were able to encode lists however they chose. In Experiment 4 the intermixing of recall and reconstruction trials in a post-cued procedure ensured that item encoding and maintenance must be the same for both tasks. Robust effects of word frequency were found in both cases. Indeed, the comparability of performance in the reconstruction tasks in Experiments 3 and 4 suggests strongly that common processes were being tapped in these two experiments.

In contrast to the inconsistent findings of word frequency effects in order reconstruction in the literature, the effects of word frequency in the current experiments are robust and orderly. When immediate tests of order memory for short lists (equivalent in length to typical serial recall tasks) were used here, the pattern is clear and more readily explained. The previous apparently 'relevant' studies have used longer presentation times, longer retention intervals, and more complicated methodologies. As a consequence, a full understanding of those data can only be gained by further research. It is, therefore, most sensible to limit the following discussion to consideration of the data reported here.

The absence of a set size effect in serial order reconstruction suggests very strongly that the set size effect seen in serial recall is one related to the retrieval of item information during output. However, we do not wish to claim that effects in the order reconstruction task cannot provide insights into processes concerning item retrieval. Despite item information being re-presented at retrieval, performance in serial reconstruction is not perfect, because the processes used to produce the next item in sequence (i.e., the same processes that drive serial recall), sometimes fail (see Neath, 1997, for more on this). We suggest that these processes are not open to introspection, although the result of the process is. Sometimes the result is the correct item: Sometimes it is an incorrect item and a number of possibilities then arise. It may be that an explicit comparison of the retrieved item against the items re-presented at test reveals to the participant which item should have been retrieved. At other times, it may be that the trace of the item is so degraded that nothing is retrieved or what does enter awareness does not discriminate between some of the items. At that point, the task demands that the participant choose a response, and errors will ensue. In agreement with Neath (1997), we conclude that even though an item is re-presented at test this does not mean that all appropriate item information will be reinstated in memory. We also accept that "reconstruction of order tests do not offer a pure measure of order memory" (Neath, 1997, p. 262).

The present experiments show a clear dissociation between set size and word frequency as manipulations of item familiarity, and, therefore, an implication is that these effects have different loci. The notion that set size influences recall through a process of redintegration during the output phase of the recall task remains a plausible explanation (as argued by Roodenrys & Quinlan, 2000). Indeed, it remains plausible that the effect of word frequency in the serial recall task also reflects characteristics of item redintegration. However, we suggest that the word frequency effects in the reconstruction tasks reflect something else.

In this regard, we agree with Thorn et al. (2009) in accepting a multiple mechanisms account in which long-term knowledge influences short-term memory in a number of different ways. We accept that processes of redintegration reflect the influence of long-term knowledge in the manner discussed (see Thorn et al., 2009), for instance, in helping reconstruct partial traces in the store during serial recall. However, it is difficult to see why such a process of reconstruction is needed when the items themselves are re-presented at test. The problem facing the system is not the reconstruction of item traces, but recovery of the item ordering from the items themselves. One possibility is that word frequency influences pre-retrieval processes as well as item retrieval. For instance, Lewandowsky and Farrell (2000) adopted a similar idea in allowing for lexical effects to emerge as a consequence of a benefit at encoding for words relative to non-words and we suggest that word frequency may play a similar role. Although they discussed effects of word frequency in terms of different mechanisms, here we claim that word frequency may exert an effect during item encoding.

Many models of serial recall incorporate distinct mechanisms for the retention of item and order information. These models generally assume that the two types of information are only recombined at output, thus leading to the prediction that serial reconstruction will be immune to item-based effects. This conclusion sits uncomfortably with the word frequency effects described here. It is possible that the relative familiarity of high frequency words vs. low frequency words means that the ordering of the particular items is more readily encoded for high vs. low frequency words.

Conclusions

In sum, the experiments allow us to conclude that word frequency effects occur in both serial recall and serial order reconstruction tasks. Collectively, the data strongly suggest that even when TBR items are re-presented at test, the recovery of their order depends on accessing both the short-term memory traces of these items and their corresponding long-term memory representations. The findings sit most comfortably with a multiple mechanisms account in which word frequency influences both retrieval and pre-retrieval processes.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64-88. doi: http://dx.doi.org/10.1016/j.jml.2006.02.002

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Brown, G. D. A., Neath, I., & Chater, N. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127-181. doi: http://dx.doi.org/10.1037//0033-295X.107.1.127

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*. doi: http://dx.doi.org/10.1037/0033-295X.114.3.539

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429-460. doi: http://dx.doi.org/10.1016/0749-596X(92)90022-P

Clarkson, L., Roodenrys, S., Miller, L. M., & Hulme, C. (2016). The phonological neighbourhood effect on short-term memory for order. *Memory, 17*, 1-12. doi: http://dx.doi.org/10.1080/09658211.2016.1179330

Cohen, D. J., Farrell, J., & Johnson, N. (2002). What very small numbers mean. *Journal of Experimental Psychology: General, 131*, 424-442. doi: http://dx.doi.org/10.1037//0096-3445.131.3.424

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1136-1146. doi: http://dx.doi.org/10.1037//0278-7393.22.5.1136

Freedman, J. L., & Loftus, E. F. (1971). Retrieval of words from long-term memory. *Journal of Verbal Learning and Verbal Behavior, 10*, 107-115. doi: http://dx.doi.org/10.1016/S0022-5371(71)80001-4

Henson, R. N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*, 73-137. doi: http://doi.org/10.1006/cogp.1998.0685

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language, 49*, 500-518. doi: http://dx.doi.org/10.1016/s0749-596x(03)00096-2

Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods, 45*, 905-927. doi: http://dx.doi.org/10.3758/s13428-013-0330-5

Lewandowsky, S., & Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality and word frequency in immediate serial recall. *Psychological Research, 63*, 163-173.

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. S. Thorn & M. P. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). Hove, UK: Psychology Press.

Merritt, P. S., DeLosh, E. L., & McDaniel, M. A. (2006). Effects of word frequency on individual-item and serial order retention: Tests of the order-encoding view. *Memory & Cognition, 34*, 1615-11627. doi: http://dx.doi.org/10.3758/BF03195924

Miller, L. M., & Roodenrys, S. (2012). Serial recall, word frequency and mixed Lists: The influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1731-1740. doi: http://dx.doi.org/10.1037/a0028470

Miller, L. M., Roodenrys, S., & Arcioni, B. (under review). *Co-articulatory fluency does not explain lexicality and frequency effects in serial recall.*

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1353-1380. doi: http://dx.doi.org/10.1037/0278-7393.29.6.1353

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition, 25*, 256-263. doi: http://dx.doi.org/10.3758/BF03201116

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology, 48A*, 384-404. doi: http://dx.doi.org/10.1080/14640749508401396

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology, 50*, 408-412. doi: http://dx.doi.org/10.1080/14640749508401396

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1019–1034. doi: http://dx.doi.org/10.1037//0278-7393.28.6.1019

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 8*, 71-78. doi: http://dx.doi.org/10.1080/096582100387623

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes Factors for ANOVA Designs. . *Journal of Mathematical Psychology, 56*, 356-374. doi: http://dx.doi.org/doi:10.1016/j.jmp.2012.08.001

Stuart, G., & Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 796–802. doi: http://dx.doi.org/10.1037/0278-7393.26.3.796

Team, J. (2017). JASP (Version 0.8.1.0).

Thorn, A. S., Frankish, C. R., & Gathercole, S. E. (2009). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. S. Thorn & M. P. A. Page (Eds.), *Interactions between Short-term and Long-term Memory in the Verbal Domain* (pp. 198-219). Hove, UK: Psychology Press.

Wetherick, N. E. (1975). The role of semantic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 471-480. doi: http://dx.doi.org/10.1016/S0022-5371(75)80025-9

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives in Psychological Science, 6*, 291-298. doi: http://dx.doi.org/10.1177/1745691611406923

Whiteman, H. L., Nairne, J. S., & Serra, M. (1994). Recognition and recall-like processes in the long-term reconstruction of order. *Memory, 2*, 275-294. doi: http://dx.doi.org/10.1080/09658219408258949

Appendix A

Stimuli used in Experiments 1 & 2

					Phonological similarity†		
Condition	Item	log(Freq.)*	PNS.	Conc.*	Onset	Nucleus	Coda
LF	babe	0.30	19	562	0.18	0.25	0.17
	barb	0.30	21	527	0.18	0.34	0.17
	barn	1.08	33	614	0.18	0.34	0.21
	bet	1.11	37	403	0.18	0.26	0.17
	bib	0.30	21	548	0.18	0.29	0.17
	bin	0.78	39	598	0.18	0.29	0.21
	bite	1.15	43	509	0.18	0.25	0.17
	bud	0.85	40	549	0.18	0.28	0.16
	cane	1.04	47	590	0.33	0.25	0.21
	cape	1.20	24	581	0.33	0.25	0.18
	carp	0.48	20	613	0.33	0.34	0.18
	cart	1.08	32	576	0.33	0.34	0.17
	coil	0.85	26	490	0.33	0.27	0.23
	cone	0.70	39	573	0.33	0.31	0.21
	cork	0.78	34	608	0.33	0.32	0.35
	cowl	0.00	23	456	0.33	0.30	0.23
	dame	0.60	25	528	0.20	0.25	0.22
	deed	1.00	31	410	0.20	0.33	0.16
	dell	0.30	27	513	0.20	0.26	0.23
	dim	1.23	28	402	0.20	0.29	0.22
	dime	0.70	28	582	0.20	0.25	0.22
	done	1.00	44	217	0.20	0.35	0.21
	dot	1.04	29	530	0.20	0.35	0.17
	dumb	1.04	35	340	0.20	0.28	0.22
	fawn	0.30	41	581	0.25	0.32	0.21
	fell	0.95	30	407	0.25	0.26	0.23
	foal	0.30	39	420	0.25	0.31	0.23
	foil	0.60	21	509	0.25	0.27	0.23
	hawk	0.85	30	623	0.29	0.32	0.35
	haze	0.78	53	509	0.29	0.25	0.24
	hide	0.70	42	451	0.29	0.25	0.16
	hood	0.78	26	547	0.29	0.35	0.16
	hop	1.00	24	494	0.29	0.35	0.18

					Phonological similarity†		
Condition	Item	log(Freq.)*	PNS.	Conc.*	Onset	Nucleus	Coda
LF *cont.*	hose	0.60	49	596	0.29	0.31	0.26
	howl	0.90	26	434	0.29	0.30	0.23
	keel	0.30	37	515	0.33	0.33	0.23
	kite	0.70	35	592	0.33	0.25	0.17
	knoll	0.30	36	486	0.27	0.35	0.23
	lace	1.15	30	545	0.28	0.25	0.26
	lard	0.30	44	517	0.28	0.34	0.16
	lark	0.60	27	578	0.28	0.34	0.35
	lease	0.78	30	371	0.28	0.33	0.26
	lice	0.48	25	543	0.28	0.25	0.26
	lime	0.95	26	590	0.28	0.25	0.22
	mall	1.04	34	417	0.25	0.32	0.23
	mat	1.15	41	513	0.25	0.29	0.17
	moat	0.60	32	453	0.25	0.31	0.17
	mole	0.78	45	590	0.25	0.31	0.23
	moss	0.85	28	575	0.25	0.35	0.26
	nip	0.30	25	515	0.27	0.29	0.18
	noose	0.00	16	542	0.27	0.39	0.26
	noun	0.30	11	387	0.27	0.30	0.21
	numb	0.70	20	379	0.27	0.28	0.22
	pall	0.30	42	362	0.19	0.32	0.23
	pat	0.95	35	400	0.19	0.29	0.17
	pearl	1.08	31	597	0.19	0.29	0.23
	peck	0.60	28	432	0.19	0.26	0.35
	peep	0.48	40	388	0.19	0.33	0.18
	pep	0.00	15	314	0.19	0.26	0.18
	pine	1.23	34	592	0.19	0.25	0.21
	poll	1.56	31	515	0.19	0.35	0.23
	pope	0.78	25	593	0.19	0.31	0.18
	puck	0.00	30	472	0.19	0.28	0.35
	pup	0.00	20	544	0.19	0.28	0.18
	rack	1.04	39	535	0.22	0.29	0.35
	rake	0.30	39	597	0.22	0.25	0.35
	ram	0.78	36	541	0.22	0.29	0.22
	rap	1.00	31	452	0.22	0.29	0.18
	reap	0.30	32	373	0.22	0.33	0.18

					Phonological similarity†		
Condition	Item	log(Freq.)*	PNS.	Conc.*	Onset	Nucleus	Coda
LF *cont.*	rhyme	0.70	31	434	0.22	0.25	0.22
	rim	0.95	31	511	0.22	0.29	0.22
	ripe	0.95	27	360	0.22	0.25	0.18
	rum	0.78	32	600	0.22	0.28	0.22
	sane	0.90	45	290	0.26	0.25	0.21
	sap	0.30	28	540	0.26	0.29	0.18
	sod	0.60	36	569	0.26	0.35	0.16
	sop	0.00	25	373	0.26	0.35	0.18
	tame	0.70	24	335	0.21	0.25	0.22
	toad	0.60	37	568	0.21	0.31	0.16
	toil	0.48	17	386	0.21	0.27	0.23
	toll	0.95	40	424	0.21	0.35	0.23
	veal	0.70	24	528	0.24	0.33	0.23
	veil	1.20	33	537	0.24	0.25	0.23
	vile	0.60	23	379	0.24	0.25	0.23
	wad	0.60	41	479	0.34	0.35	0.16
	weep	0.60	30	439	0.34	0.33	0.18
	weird	0.85	32	253	0.34	0.30	0.16
	whack	0.00	29	409	0.34	0.29	0.35
	whale	1.30	50	533	0.34	0.25	0.23
	whiff	0.48	30	413	0.34	0.29	0.27
	whip	1.18	39	570	0.34	0.29	0.18
	whirl	0.30	30	402	0.34	0.29	0.23
	whoop	0.00	18	383	0.34	0.35	0.18
	worm	1.23	19	611	0.34	0.29	0.22
	wreck	0.95	32	505	0.22	0.26	0.35
	wren	0.70	33	629	0.22	0.26	0.21
	M	0.70	31.38	491.30	0.25	0.30	0.22
	SD	0.36	8.36	92.80	0.05	0.03	0.05

Note. log(Freq.) – log (base 10) of word frequency; Conc. – concreteness, and PNS – phonological neighbourhood size.

* Frequency values were adjusted for the effects of homophones. Concreteness values are the weighted averages by frequency count across homophones.

† Phonological similarity measures using the algorithm of Mueller, Seymour, Kieras & Meyer (2003)

					Phonological similarity†		
Condition	Item	log(Freq.)*	PNS.	Conc.*	Onset	Nucleus	Coda
HF	ball	2.06	40	611	0.19	0.33	0.24
	base	1.94	26	448	0.19	0.25	0.24
	beach	1.96	21	592	0.19	0.30	0.38
	bed	2.43	44	635	0.19	0.27	0.15
	bill	1.88	39	528	0.19	0.29	0.24
	bird	2.01	46	602	0.19	0.27	0.15
	board	2.03	61	565	0.19	0.33	0.15
	boat	1.89	35	637	0.19	0.30	0.16
	bone	1.85	42	588	0.19	0.30	0.20
	book	2.64	22	609	0.19	0.35	0.34
	card	1.85	42	565	0.34	0.33	0.15
	case	2.69	26	548	0.34	0.25	0.24
	cup	1.89	19	539	0.34	0.31	0.19
	cut	1.92	30	430	0.34	0.31	0.16
	dark	2.29	21	497	0.20	0.33	0.34
	date	1.89	28	514	0.20	0.25	0.16
	dead	2.26	28	429	0.20	0.27	0.15
	deal	2.29	30	342	0.20	0.30	0.24
	farm	1.95	14	565	0.25	0.33	0.23
	fat	1.96	31	540	0.25	0.32	0.16
	feel	2.48	36	324	0.25	0.30	0.24
	feet	2.53	28	636	0.25	0.30	0.16
	fight	2.01	39	455	0.25	0.25	0.16
	firm	1.99	17	400	0.25	0.27	0.23
	form	2.55	22	438	0.25	0.33	0.23
	full	2.44	20	378	0.25	0.35	0.24
	girl	2.64	22	607	0.33	0.27	0.24
	gun	1.99	29	612	0.33	0.31	0.20
	hall	2.15	39	555	0.29	0.33	0.24
	hard	2.48	45	425	0.29	0.33	0.15
	head	2.49	38	603	0.29	0.27	0.15
	heart	2.21	28	605	0.29	0.33	0.16
	heat	2.09	31	472	0.29	0.30	0.16
	hell	1.97	33	355	0.29	0.27	0.24

					Phonological similarity†		
Condition	Item	log(Freq.)*	PNS.	Conc.*	Onset	Nucleus	Coda
HF *cont.*	hill	2.07	39	588	0.29	0.29	0.24
	hope	2.21	25	261	0.29	0.30	0.19
	hot	2.16	31	507	0.29	0.38	0.16
	job	2.52	24	432	0.40	0.38	0.18
	keep	2.36	29	339	0.34	0.30	0.19
	kid	1.94	32	536	0.34	0.29	0.15
	lead	2.27	54	543	0.28	0.30	0.15
	learn	2.10	19	370	0.28	0.27	0.20
	leg	2.24	15	626	0.28	0.27	0.33
	light	2.56	40	550	0.28	0.25	0.16
	line	2.47	40	477	0.28	0.25	0.20
	lip	1.89	27	590	0.28	0.29	0.19
	loss	1.99	25	313	0.28	0.38	0.24
	male	2.11	45	552	0.25	0.25	0.24
	mark	1.86	29	464	0.25	0.33	0.34
	mass	2.05	29	397	0.25	0.32	0.24
	meal	1.96	37	602	0.25	0.30	0.24
	meet	2.38	32	417	0.25	0.30	0.16
	mile	2.24	30	460	0.25	0.25	0.24
	mine	2.02	35	452	0.25	0.25	0.20
	mouth	2.17	9	568	0.25	0.29	0.19
	name	2.54	20	405	0.26	0.25	0.23
	neck	1.90	21	587	0.26	0.27	0.34
	nice	2.18	17	279	0.26	0.25	0.24
	night	2.68	37	498	0.26	0.25	0.16
	nine	1.87	30	452	0.26	0.25	0.20
	nose	1.91	38	628	0.26	0.30	0.23
	paid	2.09	38	386	0.20	0.25	0.15
	park	1.89	36	579	0.20	0.33	0.34
	pass	2.08	21	385	0.20	0.33	0.24
	peace	2.42	29	359	0.20	0.30	0.24
	phone	1.86	32	624	0.20	0.30	0.20
	rain	1.99	45	566	0.22	0.25	0.20
	red	2.21	36	501	0.22	0.27	0.15
	road	2.07	50	583	0.22	0.30	0.15

					Phonological similarity†		
Condition	Item	log(Freq.)*	PNS.	Conc.*	Onset	Nucleus	Coda
HF *cont.*	rock	2.39	33	600	0.22	0.38	0.34
	role	2.08	46	354	0.22	0.30	0.24
	room	2.21	31	566	0.22	0.35	0.23
	rule	2.73	30	286	0.22	0.35	0.24
	seat	2.11	49	568	0.26	0.30	0.16
	sign	2.04	36	516	0.26	0.25	0.20
	soon	2.16	25	261	0.26	0.35	0.20
	sun	2.50	37	617	0.26	0.31	0.20
	take	2.18	28	332	0.21	0.25	0.34
	talk	2.50	30	422	0.21	0.33	0.34
	team	2.11	27	489	0.21	0.30	0.23
	tell	2.01	26	306	0.21	0.27	0.24
	term	2.40	21	374	0.21	0.27	0.23
	top	2.41	27	435	0.21	0.38	0.19
	town	2.39	17	556	0.21	0.29	0.20
	turn	2.35	21	359	0.21	0.27	0.20
	type	2.25	21	376	0.21	0.25	0.19
	walk	2.15	30	452	0.35	0.33	0.34
	wall	1.91	41	589	0.35	0.33	0.24
	week	2.32	33	379	0.35	0.30	0.34
	weight	2.69	40	412	0.35	0.25	0.16
	white	2.19	50	472	0.35	0.25	0.16
	wide	2.59	49	348	0.35	0.25	0.15
	wife	2.13	26	562	0.35	0.25	0.27
	wine	2.39	45	581	0.35	0.25	0.20
	wood	1.89	25	249	0.35	0.35	0.15
	write	2.92	43	377	0.22	0.25	0.16
	M	2.21	31.82	482.96	0.26	0.29	0.22
	SD	0.27	9.79	106.96	0.05	0.04	0.06

Note. log(Freq.) – log (base 10) of word frequency; Conc. – concreteness, and PNS – phonological neighborhood size.

* Frequency values were adjusted for the effects of homophones. Concreteness values are the weighted averages by frequency count across homophones.

† Phonological similarity measures using the algorithm of Mueller, et al. (2003)

Stimuli used in Experiments 3 & 4

Condition	Item	log(Freq.)	PNS.	Conc.	Vowel length
Low	bud	0.85	40	549	1
	beak	0.85	39	552	2
	deed	1.00	31	410	2
	dot	1.04	29	530	1
	fell	0.95	30	407	1
	hide	0.70	42	451	3
	hood	0.78	26	547	1
	howl	0.90	26	434	3
	carp	0.48	20	613	2
	kite	0.70	35	592	3
	gown	1.08	9	586	3
	lice	0.48	25	543	3
	wren	0.70	33	629	1
	meek	0.30	38	299	2
	moth	0.78	14	550	1
	numb	0.70	20	379	1
	noose	0.00	16	542	2
	puck	0.00	30	472	1
	dell	0.30	27	513	1
	rake	0.30	39	597	3
	rhyme	0.70	31	434	3
	whack	0.00	29	409	1
	tuck	0.30	35	437	1
	tack	0.48	40	565	1
	toll	0.95	40	424	1
	tame	0.70	24	335	3
	wad	0.60	41	479	1
	wharf	0.48	15	573	2
	fawn	0.30	41	581	2
	whirl	0.30	30	402	2
	M	0.59	29.83	494.47	1.8
	SD	0.31	8.93	86.29	0.83

High	bird	2.01	46	602	2
	book	2.64	22	609	1
	dead	2.26	28	429	1
	date	1.89	28	514	3
	full	2.44	20	378	3
	head	2.49	38	603	1
	hard	2.48	45	425	2
	hall	2.12	39	565	2
	keep	2.36	29	339	2
	cut	1.92	30	430	1
	gun	1.99	29	612	1
	loss	1.99	25	313	1
	rain	1.91	45	600	3
	mark	1.86	29	464	2
	mouth	2.17	9	568	3
	name	2.54	20	405	3
	nice	2.18	17	279	3
	park	1.89	36	579	2
	deal	2.29	30	342	2
	rock	2.08	33	600	1
	room	2.73	31	566	2
	week	2.63	33	383	2
	talk	2.11	30	422	2
	take	2.50	28	332	3
	tell	2.40	26	306	1
	team	2.00	27	492	2
	wide	2.13	49	348	3
	wife	2.39	26	562	3
	phone	1.86	32	624	3
	wall	2.32	41	589	2
	M	2.219969	30.7	476	2.0666667
	SD	0.2575905	8.817218	112.42894	0.7717225

Appendix B

Table B1

Summary descriptives for the conditions of interest in Experiment 1

Group	List Position											
	1		2		3		4		5		6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	High Frequency											
Open	0.92	0.09	0.88	0.10	0.81	0.14	0.74	0.16	0.58	0.24	0.61	0.27
Closed	0.91	0.12	0.80	0.15	0.72	0.20	0.62	0.20	0.57	0.23	0.68	0.20
	Low Frequency											
Open	0.82	0.14	0.68	0.21	0.61	0.17	0.49	0.19	0.32	0.21	0.40	0.24
Closed	0.85	0.17	0.79	0.18	0.73	0.19	0.61	0.20	0.52	0.22	0.59	0.23

Table B2

Summary descriptives for the conditions of interest in Experiment 2

Group	List Position											
	1		2		3		4		5		6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	High Frequency											
Open	0.92	0.08	0.83	0.14	0.76	0.21	0.71	0.21	0.70	0.20	0.80	0.18
Closed	0.88	0.14	0.79	0.19	0.72	0.18	0.66	0.21	0.63	0.19	0.69	0.20
	Low Frequency											
Open	0.90	0.13	0.74	0.17	0.71	0.19	0.61	0.23	0.58	0.22	0.64	0.20
Closed	0.88	0.15	0.76	0.16	0.68	0.20	0.63	0.23	0.55	0.22	0.65	0.18

Table B3

Summary descriptives for the conditions of interest in Experiment 3

Word Frequency	List Position											
	1		2		3		4		5		6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
High	0.94	0.10	0.87	0.07	0.81	0.08	0.76	0.07	0.73	0.08	0.78	0.08
Low	0.90	0.10	0.79	0.08	0.70	0.10	0.63	0.10	0.60	0.10	0.67	0.09

Table B4

Summary descriptives for the conditions of interest in Experiment 4

Word Frequency	List Position											
	1		2		3		4		5		6	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	Recall											
High	0.86	0.12	0.79	0.16	0.71	0.18	0.62	0.18	0.53	0.19	0.55	0.22
Low	0.70	0.19	0.55	0.20	0.44	0.19	0.39	0.19	0.35	0.18	0.43	0.20
	Reconstruction											
High	0.08	0.87	0.10	0.78	0.15	0.72	0.17	0.70	0.17	0.77	0.16	0.08
Low	0.10	0.78	0.17	0.66	0.18	0.60	0.17	0.55	0.17	0.65	0.17	0.10